Signal Processing 120 (2016) 562–571

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Mitigation of pulse-width-modulation distortion using a digital predistorter based on memory polynomials

Fernando Chierchie [a,*], Juan Cousseau [a], Eduardo Paolini [a,b], Stefan Werner [c]

[a] Inst. de Investigaciones en Ing. Eléctrica - Alfredo Desages (UNS-CONICET), Dto. de Ing. Eléctrica y de Computadoras, Universidad Nacional del Sur, Av. Alem 1253 (8000), Bahía Blanca, Argentina
[b] Comisión de Investigaciones Científicas (CIC), Buenos Aires, Argentina
[c] Department of Signal Processing and Acoustics, Aalto University, School of Electrical Engineering, P.O. Box 13000, 00076 Aalto, Finland

ARTICLE INFO

Article history:
Received 7 April 2015
Received in revised form
24 September 2015
Accepted 17 October 2015
Available online 26 October 2015

Keywords:
Pulse width modulation
Baseband
Digital predistorter
Indirect learning

ABSTRACT

This paper presents an adaptive digital predistorter (DPD) that significantly reduces the baseband distortion of digital pulse width modulation (PWM), typically used in switching (class-D) amplifiers. A generalized Hammerstein structure (also known as power filter) composed by static nonlinearities and FIR filters is used to model the baseband behavior of PWM. We show that the contribution of the higher order terms of the nonlinearity are negligible and therefore, for practical applications, only the first three or four odd powers should be retained. The convergence of the DPD is studied and the performance is demonstrated and compared with other approaches. Good results are obtained for typical signals even when using low-order FIR filters in the DPD. Measurement results obtained using a digital signal processor are also presented aiming to validate the proposed approach.

© 2015 Published by Elsevier B.V.

1. Introduction

Pulse width modulation (PWM) is a time-encoding technique which allows to modulate an arbitrary band-limited signal into a binary signal with pulses having variable widths. This binary signal can be used to drive amplifier circuits that have transistors operating as switches (on-off, rather than in their active region), which results in a reduced power consumption, typically achieving an efficiency above 90%. PWM is used in different applications, from power electronics [1], to audio amplifiers [2] and recently, RF transmitters [3]. It is particularly relevant for portable applications [4], where efficiency is one of the main concerns, e.g. hearing aids [5,6] that require high efficiency and low distortion. Usually, the amplified version of the modulating signal is recovered (demodulated) in the analog domain using a low-pass filter composed of passive elements [7]. Any baseband distortion produced by the modulation process will appear in the demodulated signal since the passive linear filter cannot distinguish signal from baseband distortion.

Although there are various PWM schemes, ranging from analog PWM, or natural PWM (NPWM), to discrete-time implementations, where the most commonly known is uniform PWM (UPWM) [8], all PWM methods introduce baseband distortion. In NPWM, the distortion is composed of sidebands of the carrier frequency, and for practical applications it can be reduced to negligible levels by increasing the carrier frequency, but at the expense of a decreased efficiency. UPWM exhibits greater harmonic distortion than NPWM, and the distortion is not only composed by the sidebands of the carrier but also by harmonics of the modulating signal. Therefore, UPWM distortion cannot be efficiently reduced by increasing the carrier frequency [9].

* Corresponding author.
E-mail address: fernando.chierchie@uns.edu.ar (F. Chierchie).

http://dx.doi.org/10.1016/j.sigpro.2015.10.017
0165-1684/© 2015 Published by Elsevier B.V.

In the majority of current applications, the modulating signal is available as discrete-time samples, hence digital PWM is preferred. Research efforts were initially devoted to obtain digital PWM modulation alternatives with reduced distortion within the baseband. These efforts were focused on imitating the NPWM behavior through digital signal processing and resulted in a method usually known as pseudo NPWM [10,11] or interpolated PWM, which involves interpolation between discrete samples and a crossing-point estimation between the interpolated signal and the carrier signal. A statistically optimal estimator for interpolated PWM was reported in [12]. At their best, these methods can obtain the spectral performance of NPWM. A different approach is to design a digital modulation algorithm without distortion components in the baseband. One of the first attempts was Click modulation [13,14] which is based on properties of analytic signals. This complex modulation scheme is capable of representing band-limited signals using PWM-like waveforms with zero distortion in the baseband. However, its implementation in real time is still computationally intensive.

Volterra type models have been widely used for modelling and design of digital predistortion (DPD) compensators for radio frequency (RF) power amplifiers [15–18], which are typically class-AB amplifiers but they are less common for switched amplifiers (class-D). The DPD can be obtained as a pth-order inverse of the model as in [19]. In general development of the inverse system is not straightforward and the complexity grows rapidly with the order-p [20,21,15].

In this work, first a parallel Hammerstein model of the PWM process (also known as power filter [22]) composed of the parallel connection of an static nonlinearity followed by a discrete-time FIR filter is obtained. It is shown that the contribution of the higher-order powers are bounded to negligible levels and, therefore, for practical applications, only the first three or four odd powers are required to obtain a reliable model. Secondly, a parallel Hammerstein structure using FIR filters is used for digital predistortion. The coefficients of the DPD are obtained adaptively, requiring a small number of parameters to reduce the distortion of the PWM to negligible levels. For the adaptive DPD, the identification problem is linear in the parameters and the recursive least-squares algorithm (RLS) is used offline to adapt the coefficients of the DPD at the training stage. The properties of the mean-square error convergence of the weights are analyzed. Although the theoretical results are derived assuming independent identically distributed (iid) signals, simulation results show that the proposed DPD achieves a significant improvement in the signal-to-noise ratio (SNR) also for non-iid inputs. The performance of the DPD is also tested using an *indirect learning architecture* (ILA) where the parameters are computed online [15,20]. Although the computational effort is higher, the added benefits are that exact knowledge of the nonlinear amplifier model is not required and better SNR levels are achieved.

The reduction of baseband distortion allows for the use of PWM amplifiers with a low ratio between the carrier frequency and the maximum frequency of the modulating signal which contributes to efficiency. The proposed DPD shows a good performance as compared to other alternatives even when relatively short-length FIR filters (with impulse responses shorter than 40 samples) are used, reducing the computational complexity and latency introduced by the DPD.

The organization of the paper is as follows. The discrete-time parallel Hammerstein model of the PWM is reviewed in Section 2 and a bound to the contribution of each power of the nonlinearity is derived. In Section 3, the DPD architecture is discussed and the convergence in the mean-square sense is studied. Simulation results, comparisons to previous results and experimental validation are shown in Section 4. Finally, conclusions are drawn in Section 5.

2. Discrete-time baseband model for PWM

A PWM signal is composed of pulses whose widths depend on the modulating signal. In this presentation it will be assumed that the distance T between the center of the pulses is fixed, and it will be referred as the PWM period (symmetric PWM). The inverse of the PWM period is the carrier frequency or PWM frequency $f_s = 1/T$. The PWM signal can be represented as [9,23]

$$p(t) = \sum_{n=-\infty}^{\infty} \left[u\left(t - nT + w_n \frac{T}{2}\right) - u\left(t - nT - w_n \frac{T}{2}\right) \right] \tag{1}$$

where $u(t)$ is the Heaviside step function ($u(t) = 0$ if $t < 0$; $u(t) = 1$ if $t \geq 0$), and $0 \leq w_n < 1$ is the n-th normalized duty cycle. Eq. (1) represents a two level PWM signal, *i.e.* a binary signal with only two possible amplitude values 0 and 1. In digital PWM, the duty cycles w_n usually exhibit an affine dependence on the samples of the modulating signal: $x_n = x(t)|_{t=nT}$.[1]

To retain the baseband characteristic of the PWM modulation, $p(t)$ is filtered with an ideal low pass filter. That is, the baseband signal is obtained as $y(t) = h_{LP}(t) * p(t)$, where $h_{LP}(t) = f_s \mathrm{sinc}(f_s t)$ is the impulse response of the low pass filter with cut-off frequency $f_s/2$ and "$*$" denotes convolution. Uniform sampling of $y(t)$ at f_s yields [19,23]

$$y_n = f_0(w_n) + \sum_{k=1}^{\infty} [f_k(w_{n-k}) + f_k(w_{n+k})] = \sum_{k=-\infty}^{\infty} f_k(w_{n-k}) \tag{2}$$

with

$$f_k(w) \triangleq \left(\mathrm{Si}\left[k\pi + w(\pi/2)\right] - \mathrm{Si}\left[k\pi - w(\pi/2)\right]\right)/\pi, \tag{3}$$

where $\mathrm{Si}(z) = \int_0^z \sin(\tau)/\tau \, d\tau$ is the sine integral function.

In summary, the modulating signal samples x_n are used to derive the duty cycles w_n, which in turn allows the construction of the analog PWM signal $p(t)$ in (1). The baseband content in the frequency range $[0 \; f_s/2]$ of $p(t)$ is represented by the discrete-time samples y_n in (2). We call this model the "PWM baseband model" since it replicates

[1] Typically, in UPWM the duty cycles are computed as an scaled version of the input samples or $w_n = x_n$ if $0 \leq x_n < 1$. This direct signal-amplitude to pulse-width conversion introduces some degree of distortion which cannot be eliminated by the low-pass demodulator filter.

Table 1
Impulse responses $g_{i,n}$ for the nonlinear PWM model (5).

Sample index	$g_{1,n}$	$g_{3,n}$	$g_{5,n}$	$g_{7,n}$
$n=0$	$b_0=1$	$b_1=-\pi^2/72$	$b_2=\pi^4/9600$	$b_3=-\pi^6/2257920$
$n\neq 0$	$c_0=0$	$c_1=\frac{(-1)^{n+1}}{12n^2}$	$c_2=\frac{(-1)^n(n^2\pi^2-6)}{480n^4}$	$c_3=\frac{(-1)^n(20n^2\pi^2-n^4\pi^4-120)}{53760n^6}$

the behavior of the PWM modulator in the frequency band $0\le f\le f_s/2$.

The objective of zero baseband distortion is accomplished if the duty cycles w_n are derived in such a way that $y_n\equiv x_n$, for all n, *i.e.*, when the samples of the low-pass filtered PWM signal are equal to the samples of the modulating signal.

2.1. Parallel Hammerstein model

Based on (2) and (3), a generalized or parallel discrete-time Hammerstein model can be derived for PWM. The function $f_k(w)$ in (3) can be expanded in an odd power series as,

$$f_k(w)=\begin{cases} b_0w+b_1w^3+b_2w^5+b_3w^7+\cdots, & \text{if } k=0,\\ c_1w^3+c_2w^5+c_3w^7+\cdots, & \text{if } k\neq 0.\end{cases} \tag{4}$$

Replacing (4) in (2) and defining $g_{i,n}=b_{(i-1)/2}$ for $n=0$ and $g_{i,n}=c_{(i-1)/2}$ for $n\neq 0$, the baseband PWM samples can be computed as

$$\begin{aligned} y_n &= \sum_{k=-\infty}^{\infty}\left[g_{1,k}w_{n-k}+g_{3,k}w_{n-k}^3+g_{5,k}w_{n-k}^5+g_{7,k}w_{n-k}^7+\cdots\right]\\ &= g_{1,n}*w_n+g_{3,n}*w_n^3+g_{5,n}*w_n^5+g_{7,n}*w_n^7+\cdots \end{aligned} \tag{5}$$

which results in the computation of several discrete-time convolutions between the i-th power of the duty cycle signal w_n^i and the filter with impulse response $g_{i,n}$. Eq. (5) represents a simplified Volterra-type system with no cross terms [24]. As shown later, an accurate representation can be obtained using just a few low-order power terms in (5). Explicit expressions for impulse responses of order 1, 3, 5 and 7 are described in Table 1.

If the impulse responses described in Table 1 are truncated to $\hat{M}$ terms, each $g_{i,n}*w_n^i$ term can be thought as the output of a FIR filter and the output samples can be computed as

$$\begin{aligned} y_n &= \sum_{k=0}^{\hat{M}-1}\left[g_{1,k}w_{n-k}+g_{3,k}w_{n-k}^3+g_{5,k}w_{n-k}^5+g_{7,k}w_{n-k}^7+\cdots\right]\\ &= y_{1,n}+y_{3,n}+y_{5,n}+y_{7,n}+\cdots. \end{aligned} \tag{6}$$

A block diagram of such a system is shown in Fig. 1. The branches of the model are composed of a static nonlinearity followed by an FIR filter that incorporates the system memory, *i.e.*, each branch contains a Hammerstein system. The magnitude of the output samples $y_{i,n}$ of each branch are bounded as stated by the following lemma.

Lemma 1. *The contribution $y_{i,n}$ of the i-th power in (6) to the output y_n is bounded by*:

$$y_{i,n}\le (B_w)^i|G_i(e^{j\pi})|$$

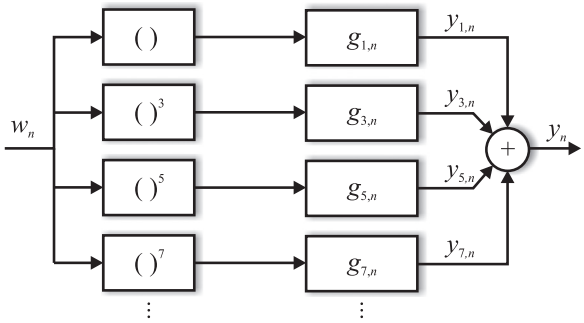

Fig. 1. Parallel Hammerstein model for PWM.

where B_w is the bound on the duty cycles $w_n\le B_w<1$, and $|G_i(e^{j\pi})|$ is the magnitude of the frequency response $G_i(e^{j\omega})$ of the filter $g_{i,n}$ evaluated at $\omega=\pi$.

Proof of Lemma 1. The contribution of the i-th power can be bounded using the convolution sum,

$$\begin{aligned} |y_{i,n}| &= |\sum_k g_{i,k}(w_{n-k})^i|\le \sum_k |g_{i,k}||(w_{n-k})^i|\le (B_w)^i\sum_k|g_{i,k}|\\ &= (B_w)^i|G_i(e^{j\pi})|. \end{aligned} \tag{7}$$

The last identity is based on the discrete-time Fourier transform definition and the alternating sign of $g_{i,n}$ due to the $(-1)^n$ factor (see Table 1): $|G_i(e^{j\pi})|=|\sum_k g_{i,k}(-1)^k|=\sum_k|g_{i,k}|$. This completes the proof.□

Remark. According to Lemma 1, the i-th branch of model (6) (see Fig. 1) for $i>7$ produces a negligible contribution $y_{i,n}$ to y_n, because $w_n\le B_w<1$ and hence $(w_n)^i\ll 1$. Furthermore, the associated filter with impulse response $g_{i,k}$ not only blocks the DC-component for $i>1$, but also increasingly attenuates $(w_n)^i$ as i grows larger, as depicted in Fig. 2 where the amplitude $|G_i(e^{j\omega})|$ of the frequency responses of the filters are shown. If $i=9$ and $B_w=0.6$ (see sampling theorem for PWM signals in [23]) then, $|G_9(e^{j\pi})|\approx 1\times 10^{-4}$ and $(w_n)^9<0.01$. Hence, using Lemma 1, one notices that the contribution of the 9-th power branch is less than $|y_{i,n}|\le 1\times 10^{-6}$ (less than -120 dB). From the high-pass characteristic of the frequency responses shown in Fig. 2 it can be inferred that if the signal bandwidth is reduced, or if the signals are of low-pass nature (which is usually the case for audio signals), the contribution of the higher order powers will be further reduced.

3. Adaptive predistortion based on the parallel Hammerstein structure

A widespread architecture for digital predistortion is the indirect learning architecture (ILA) [20]; a block diagram is shown in Fig. 3. In this scheme the DPD should approximate the inverse of the system (in our case the PWM baseband model), producing from the input signal x_n a signal w_n which is used as input to the PWM block to generate the output $y_n \approx x_n$ [25]. As shown in the block diagram of Fig. 3 a predistorter training (A) is used to compute the coefficients of the DPD and then these coefficients are used in the actual DPD copy of (A). The output of the system, y_n, is used as the input to the adaptive DPD training block and then the output z_n is compared with PWM input w_n. If the training DPD approximates the inverse of the nonlinear system, then the difference between the PWM input and the DPD training output $\alpha(n)$ approaches zero since $z_n \approx x_n$.

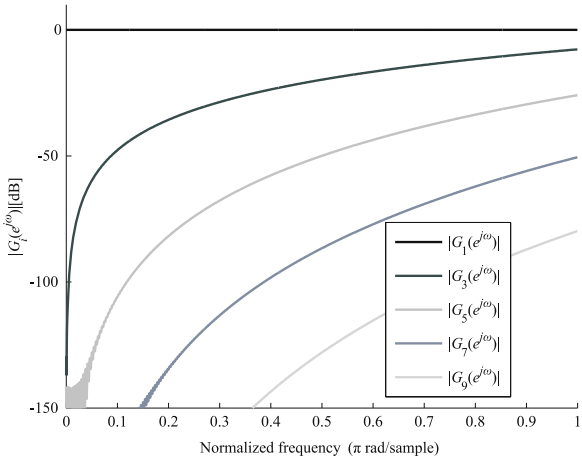

Fig. 2. Amplitude frequency responses $|G_i(e^{j\omega})|$ of the filters with impulse responses $g_{i,n}$.

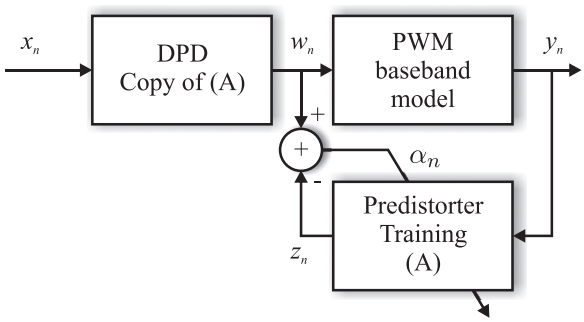

Fig. 3. DPD based in the indirect learning architecture (ILA).

For certain applications where the real-time adaptation scheme of the ILA architecture may be computationally too expensive and perhaps unnecessary, offline training of the DPD can be considered. The block diagrams of the structures used for offline training and real-time operation of the DPD are shown in Fig. 4. The DPD training is shown in Fig. 4(a), where the weights of a parallel Hammerstein structure are computed adaptively ("DPD training"). Once convergence is achieved, the DPD operates as shown in Fig. 4(b). The predistorter signal w_n at the output of the DPD is obtained from the input signal x_n and it should be such that $y_n \approx x_n$, with some possible delay due to the prefilter stage.

3.1. Post-inverse and pre-inverse of nonlinear systems

In both the ILA architecture of Fig. 3, and the offline training of Fig. 4 the DPD works as a post-inverse during the training stage and as a pre-inverse during operation. Care should be taken when considering this approach for general types of nonlinear systems since the pre-inverse and the post-inverse may be (very) different. In the pioneer work of Eun and Powers [20] authors impose a restriction to the nonlinear system such that if $x_n \neq y_n$ then, $w_n \neq z_n$ and if $x_n = y_n$ then, $w_n = z_n$. With this restriction, if $\alpha_n = w_n - z_n$ approaches zero, thus so does $y_n - x_n$. In this way the post-inverse and pre-inverse issue is avoided.

Based on the results of Eun and Powers [20], Morgan et al. [16] proposed a memory polynomial for predistortion of RF power amplifiers using the ILA architecture. Authors justify the post-filtering and pre-filtering interchange of the ILA architecture based on a result by Schetzen [26, Ch. 7] that shows that the pth-order postinverse of a general Volterra system is identical to the pth-order preinverse. In general, the aforementioned scheme will work if the input signal x_n is sufficiently close to the signal y_n that was used

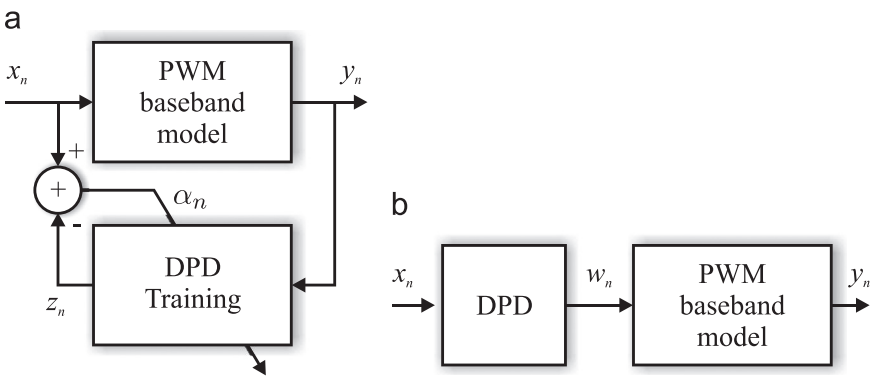

Fig. 4. DPD during training (a); DPD during operation (b).

during the offline training [25]. In the case of PWM applications, this will be satisfied under typical operating conditions, *i.e.* when the PWM carrier frequency f_s is several times higher than the maximum frequency of the input signal x_n and therefore the distortion introduced by the PWM is mild as shown by Lemma 1 and as discussed in Section 3.3.1.

One of the advantages of this technique is that it does not require an exact knowledge of the nonlinearity of the amplifier. For the PWM application, the model derived in Section 2 is fairly accurate so the DPD is expected to perform well in reducing the inherent PWM baseband distortion when a real PWM modulator is used. Both methods (ILA and offline training) have the advantage of directly estimating the inverse and thus avoids the two-step procedure of first estimating the parameters of the amplifier and, thereafter, designing the inverse [27]. The following analysis will be performed for the case shown in Fig. 4, while the ILA will be considered using simulation in Section 4.

3.2. Adaptive identification of the parallel Hammerstein structure

With N input-output data pairs, the relation between the output z_n and the input y_n of the training DPD in Fig. 4 (a) can be written in matrix form as

$$\mathbf{z} = \mathbf{Y}\mathbf{h}, \tag{8}$$

with $\mathbf{h}$, the vector of coefficients to be identified, given by

$$\mathbf{h} = \left[h_{1,0}, h_{1,1}, \ldots, h_{1,(M-1)}, h_{3,0}, h_{3,1}, \ldots, h_{3,(M-1)}, \ldots, h_{K,0}, h_{K,1}, \ldots, h_{K,(M-1)}\right]^T, \tag{9}$$

where M is the number of coefficients of the FIR filters used for the DPD, K is the maximum power used and $h_{i,j}$ is the filter coefficient that corresponds to the j-th delay of the i-th power of the input. Since only odd powers are considered, the number of coefficients to be estimated is $M \times (K+1)/2$. The vector of output samples of the adaptive filter is

$$\mathbf{z} = [z_0, z_1, \ldots, z_{N-1}]^T$$

and $\mathbf{Y}$ is the data matrix of the input y_n and its odd powers which can be written as

$$\mathbf{Y} = \begin{bmatrix} y_0 & y_{-1} & \cdots & y_{-(M-1)} & y_0^3 & y_{-1}^3 & \cdots & y_{-(M-1)}^3 & \cdots & y_0^K & y_{-1}^K & \cdots & y_{-(M-1)}^K \\ y_1 & y_0 & \cdots & y_{-(M-2)} & y_1^3 & y_0^3 & \cdots & y_{-(M-2)}^3 & \cdots & y_1^K & y_0^K & \cdots & y_{-(M-2)}^K \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots & \\ y_{N-1} & y_{N-2} & \cdots & y_{N-M} & y_{N-1}^3 & y_{N-2}^3 & \cdots & y_{N-M}^3 & \cdots & y_{N-1}^K & y_{N-2}^K & \cdots & y_{N-M}^K \end{bmatrix}$$

or, in a more compact form as

$$\mathbf{Y} = \left[\mathbf{y}_{1,0}, \mathbf{y}_{1,1}, \ldots, \mathbf{y}_{1,(M-1)}, \mathbf{y}_{3,0}, \mathbf{y}_{3,1}, \ldots, \mathbf{y}_{3,(M-1)}, \ldots, \mathbf{y}_{K,0}, \mathbf{y}_{K,1}, \ldots, \mathbf{y}_{K,(M-1)}\right] \tag{10}$$

where $\mathbf{y}_{k,m} = [y_{0-m}^k, y_{1-m}^k, \ldots, y_{N-1-m}^k]^T$. The DPD is a nonlinear adaptive filter but since the estimation error is linear in the parameters (the coefficients of the impulse responses), it is possible to derive the least-squares (LS) optimal solution $\hat{\mathbf{h}}$ as

$$\hat{\mathbf{h}} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{z}. \tag{11}$$

A stable, well-known technique to compute $\hat{\mathbf{h}}$ is the RLS algorithm [20,28] which is described by the following expressions:

$$\begin{aligned} \hat{\mathbf{h}}(n) &= \hat{\mathbf{h}}(n-1) + \mathbf{k}(n)\alpha(n). \\ \mathbf{k}(n) &= \frac{\mathbf{P}(n-1)\mathbf{Y}(n,:)}{\lambda + \mathbf{Y}^T(n,:)\mathbf{P}(n-1)\mathbf{Y}(n,:)} \\ \alpha(n) &= x_n - \hat{\mathbf{h}}(n-1)\mathbf{Y}^T(n,:) \\ \mathbf{P}(n) &= \lambda^{-1}\left[\mathbf{P}(n-1) - \mathbf{k}(n)\mathbf{Y}^T(n,:)\mathbf{P}(n-1)\right] \end{aligned} \tag{12}$$

where n is the iteration number, $0 \ll \lambda \leq 1$ is a forgetting factor, and $\mathbf{Y}(n,:)$ denotes the n-th row of $\mathbf{Y}$. The algorithm is initialized with $\mathbf{P}(0) = \delta^{-1}\mathbf{I}$, where $\delta \approx 0$ is a positive constant and all signals involved are assumed to be real-valued.

In [22] the same adaptive parallel Hammerstein structure (referred as power filter) is presented for acoustic echo cancellation. A DFT-domain implementation of the nonlinear filter as an extension of the linear overlap-save method is presented. A Gram–Schmidt based orthogonalization of the input signal x_n and its powers is also introduced. Although for faster convergence the orthogonalization may be required in certain applications [22], the convergence to the solution is relatively fast in our application and therefore the orthogonalization process was avoided. The properties of the convergence of the adaptive Hammerstein structure are analyzed in the following section.

3.3. Convergence analysis

The convergence analysis of the parallel Hammerstein structure follows the same steps as the standard recursive least-squares algorithm (RLS) [28] because the estimation error is linear in the parameters. The main differences are due to the powers of the Hammerstein structure and will be analyzed in this section. It is shown that convergence in the mean-square sense depends on the statistical distribution of the input signal, as described in [29].

The RLS based identification of the parallel Hammerstein structure is based on the following assumptions:

A1. The physical phenomena under identification responds to a model of the form

$$z_n = e_n + \mathbf{h}_0^T \mathbf{Y}(n,:) \tag{13}$$

where $\mathbf{h}_0$ is the vector of true parameters and e_n is additive noise.

A2. The noise e_n is an iid non-observable random process satisfying $\mathbf{E}[e_n]=0$ and has no correlation with any other signal in the system.

A3. The input sequence y_n of the adaptive DPD is a zero mean iid random process with arbitrary (symmetric) statistical distribution.

A4. Only odd power nonlinearities are considered for the DPD.

Assumptions A1–A3 are typically employed in the analysis of the linear RLS together with the independence assumption [28] which, among others, presumes that the tap-input vectors are statistically independent. The identification of general Hammerstein systems for non-iid inputs is still an open problem [30,29]. Regarding the zero-mean property in assumption A3, the total DC-gain of some nonlinear systems, represented using Volterra-type models, may impose an ambiguity issue because it depends on the product of the DC-gain of the static nonlinearity and the DC-gain of the linear filter representing the dynamics of the system. This is not the case for the nonlinear parallel Hammerstein model of the PWM modulator because the DC-gain of the system is constant and of known value (DC-gain=1). The mean value of the input signal is preserved by the system and passes unaltered from the input to the output. This can be verified as follows: the PWM model shown in Fig. 1 with the magnitude of the frequency responses of the filters shown in Fig. 2 reveals that the mean is preserved because the filters of order higher than one have zero DC-gain and the linear filter which corresponds to the first branch with impulse response $g_{1,n}$ (which, as shown in Table 1, is an impulse) passes the input signal unaltered to the output. This property can be also interpreted from the physical-system under analysis: $w_n = 1/2$ for all n represents a fixed duty-cycle PWM square-wave $p(t)$ with amplitude values 0 and 1 and with 50% duty cycle (1/2 mean) which after low-pass filtering and sampling yields a discrete-time signal y_n with a mean of 1/2. Since the mean value is unaffected by the nonlinear system, the DPD also preserves the mean of the signal and therefore the error signal (α_n in Fig. 4(a)) has zero mean thus not affecting the convergence properties. The iid property in A3 is discussed in detail in Section 3.3.1.

The coefficients $\mathbf{h}$ in (9) of the generalized Hammerstein system described by (8), which are adapted by the RLS algorithm defined by (12), converge in mean-square sense to the desired set of parameters $\mathbf{h}_0$ under assumptions A1–A4. The mean-square error $\mathbf{E}[||\mathbf{h}_0 - \hat{\mathbf{h}}(n)||^2]$ as a function of the autocorrelation matrix $\mathbf{R}_K = \mathbf{E}[\mathbf{Y}(n,:)\mathbf{Y}^T(n,:)]$ is [28]:

$$\xi(n) = \frac{\sigma_0^2}{n}\,\mathrm{tr}\left(\mathbf{R}_K^{-1}\right). \tag{14}$$

The properties of the convergence of (14) depend on the characteristics of the correlation matrix $\mathbf{R}_K$.

Theorem 2. *Convergence in the mean-square sense depends on the statistical distribution of the input of the Hammerstein system, particularly on the even moments up to order* 2K: $\mathbf{E}[y_n^{2K}]$.

Proof. For the parallel Hammerstein structure, the correlation matrix $\mathbf{R}_K$ of the data vector $\mathbf{Y}(n,:)$ depends not only on the tap-input vector but also on the odd powers of the input. The correlation matrix for $\mathbf{Y}(n,:)$ is a block matrix of the form

$$\mathbf{R}_K = \mathbf{E}[\mathbf{Y}(n,:)\mathbf{Y}^T(n,:)] = \mathbf{E}\begin{bmatrix} \mathbf{B}_{1,1} & \mathbf{B}_{1,2} & \cdots & \mathbf{B}_{1,q} \\ \mathbf{B}_{2,1} & \mathbf{B}_{2,2} & \cdots & \mathbf{B}_{2,q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{q,1} & \mathbf{B}_{q,2} & \cdots & \mathbf{B}_{q,q} \end{bmatrix} \tag{15}$$

where $q=(K+1)/2$ and $\mathbf{B}_{a,b}$ is a $M \times M$ matrix given by the outer product

$$\mathbf{B}_{a,b} = \begin{bmatrix} y_i^{2a-1} \\ y_{i-1}^{2a-1} \\ \vdots \\ y_{i-M+1}^{2a-1} \end{bmatrix} \begin{bmatrix} y_i^{2b-1} & y_{i-1}^{2b-1} \cdots y_{i-M+1}^{2b-1} \end{bmatrix}. \tag{16}$$

Since y_n is assumed to be an iid random process with zero mean (A3), each individual element of the matrix $\mathbf{R}_K$ falls into one of the following cases:

$$\mathbf{E}[y_i^{2a-1} y_j^{2b-1}] = \begin{cases} m_{2(a+b-1)}, & \text{if } i=j, \\ 0, & \text{if } i \neq j, \end{cases} \tag{17}$$

where $m_{2(a+b-1)}$ is the $2(a+b-1)$-th order moment of y_i, *i.e.*, $m_{2(a+b-1)} = \mathbf{E}[y_i^{2(a+b-1)}]$. Whenever $i \neq j$, y_i^{2a-1} is independent of y_j^{2b-1}. Since $2a-1$ and $2b-1$ are odd numbers, the statistical distribution is symmetric and $\mathbf{E}[y_n]=0$ then $\mathbf{E}[y_i^{2a-1}]=0$ and $\mathbf{E}[y_j^{2b-1}]=0$ and hence $\mathbf{E}[y_i^{2a-1} y_j^{2b-1}]=0$. With this simplification, the correlation matrix $\mathbf{R}_K$ can be rewritten in terms of the even moments of the input to get,

$$\mathbf{R}_K = \begin{bmatrix} m_2\mathbf{I}_M & m_4\mathbf{I}_M & \cdots & m_{K+1}\mathbf{I}_M \\ m_4\mathbf{I}_M & m_6\mathbf{I}_M & \cdots & m_{K+3}\mathbf{I}_M \\ \vdots & \vdots & \ddots & \vdots \\ m_{K+1}\mathbf{I}_M & m_{K+3}\mathbf{I}_M & \cdots & m_{2K}\mathbf{I}_M \end{bmatrix} \tag{18}$$

with $\mathbf{I}_M$ representing the $M \times M$ identity matrix. Matrix $\mathbf{R}_K$ can be rewritten using Kronecker product as

$$\mathbf{R}_K = \begin{bmatrix} m_2 & m_4 & \cdots & m_{K+1} \\ m_4 & m_6 & \cdots & m_{K+3} \\ \vdots & \vdots & \ddots & \vdots \\ m_{K+1} & m_{K+3} & \cdots & m_{2K} \end{bmatrix} \otimes \mathbf{I}_M = \mathbf{\Theta} \otimes \mathbf{I}_M \tag{19}$$

where $\mathbf{\Theta}$ is a Hankel-type matrix, *i.e.*, its skew-diagonals have constant values. Hankel matrices commonly arise in problems involving power moments [31], e.g., identification problems using Volterra type filters [32].

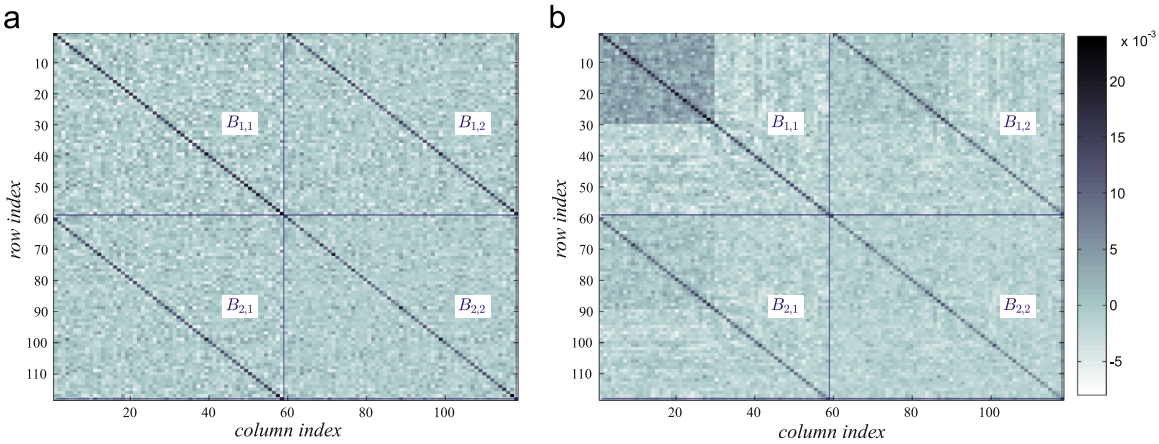

Fig. 5. Graphical representation of matrix $\mathbf{R}_K$ when: (a) y_n is pseudo-white random noise and (b) y_n is colored by the PWM model.

The mean-square error $\mathbf{E}[||\mathbf{h}_0-\hat{\mathbf{h}}(n)||^2]$ is dependent on the trace of the inverse of $\mathbf{R}_K$, and hence on the sum of the inverse of its eigenvalues:

$$\xi(n)=\mathbf{E}[||\mathbf{h}_0-\hat{\mathbf{h}}(n)||^2]=\frac{\sigma_0^2}{n}\operatorname{tr}\left(\mathbf{R}_K^{-1}\right)=\frac{\sigma_0^2}{n}\sum_{i=1}^{qM}\frac{1}{\mu_i} \tag{20}$$

where μ_i, $i=1,2,\ldots,qM$, are the eigenvalues of $\mathbf{R}_K$. The eigenvalues μ_i can be found using (19) and a known property of the Kronecker product [33]: if $\lambda_1,\lambda_2,\ldots,\lambda_q$ are the eigenvalues of $\mathbf{\Theta}$ and $\gamma_1,\gamma_2,\ldots,\gamma_M$ the eigenvalues of $\mathbf{I}_M$ then $\mu_n=\lambda_i\gamma_j$ for $i=1,\ldots,q$ and $j=1,\ldots,M$. Since $\mathbf{I}_M$ is the identity matrix $\gamma_i=1$ for all i and then the eigenvalues of $\mathbf{R}_K$ are the eigenvalues of $\mathbf{\Theta}$ each one with multiplicity M. Therefore the mean-square error is given by

$$\xi(n)=\frac{\sigma_0^2}{n}\sum_{i=1}^{q}\frac{M}{\lambda_i}, \tag{21}$$

showing that $\xi(n)\to 0$ when $n\to\infty$. The convergence is dependent on the eigenvalues λ_i of $\mathbf{\Theta}$, the matrix of moments of the input random process given by (19). This completes the proof.□

3.3.1. A comment regarding the inverse modeling problem

The inverse modelling or inverse filtering problem depicted in Fig. 4(a) is common in many applications being channel equalization (deconvolution) in communication systems a typical example. The idea is to find the best least square approximation to the inverse of the system, in our case the nonlinear PWM model. Although the input to the system x_n may be iid, the output y_n, which is the adaptive filter input is correlated by the system; in the case of the PWM model of Fig. 1 the frequency response $G_i(e^{j\omega})$ of the filters depicted in Fig. 2. Filter $g_{1,n}$ has no memory and the output sequence is uncorrelated. Certain amount of correlation is introduced by the other filters $g_{3,n}, g_{5,n}, \ldots$but, as shown by Lemma 1 this contribution is bounded. An example is shown in Fig. 5, which depicts a graphical representation of a numerical simulation of the block matrix $\mathbf{R}_K$ in (15), when y_n is a pseudo-white random Gaussian noise sequence (Fig. 5(a)) and when y_n is colored by the PWM model with parameters $K=3$ (maximum power of the model) and length of the FIRs filters $\hat{M}=59$ (Fig. 5(b)). In both cases matrices $B_{a,b}$ are diagonally dominant, although in Fig. 5(b) some correlation is observed in the upper left corner of the matrix (darker area).

When some correlation is introduced, matrix $\mathbf{R}_K$ can be thought as the sum of two matrices: $\mathbf{R}_K=\mathbf{R}_K^{iid}+\mathbf{\Delta}$, where $\mathbf{R}_K^{iid}$ is given by (19), when the input to the adaptive filter is iid and $\mathbf{\Delta}$ is a perturbation matrix with a diagonal of zeros. If the eigenvalues of $\mathbf{R}_K^{iid}$ are $v_1\geq\cdots\geq v_{qM}$ and the eigenvalues of $\mathbf{\Delta}$ are $\rho_1\cdots\geq\rho_{qM}$ then, based on Weyl's theorem [31, p. 181] the eigenvalues μ_i of $\mathbf{R}_K$ satisfy the inequality $v_i+\rho_1\leq\mu_i\leq v_i+\rho_{qM}$. Furthermore, if $\mathbf{\Delta}_{i,j}$ are small, the eigenvalues ρ_i of $\mathbf{\Delta}$ can be shown to lie on a disc of radius $||\mathbf{\Delta}||_1$ centred at zero, by the Geršgorin Disc Theorem [31, p. 344]. In other words, if the entries of the perturbation matrix are small then there will only be a small spread of the eigenvalues of $\mathbf{R}_K$ compared to those of $\mathbf{R}_K^{iid}$.

4. Simulation and validation experiments

We consider two frameworks. In the first case, "online DPD", the parameters are computed in real time based on the actual input signal as shown in Fig. 3 (ILA architecture). In the second case, ("offline DPD") the predistorter is trained offline with certain signal. The parameters of the DPD are then fixed and its performance is tested with other signals with similar statistical properties. The offline DPD is also tested using a real-time DSP-based application, and experimental laboratory measurements are shown.

4.1. Simulation scheme

To evaluate the DPD, the PWM model given by (6) with FIRs filters having impulse responses of length $\hat{M}=999$ and with nonlinearity up to order 7 was used. The sampling and PWM frequency is normalized to $f_s=1$. Different band-limited signals (colored noise and music) were used to test and compare the algorithms. Signal A

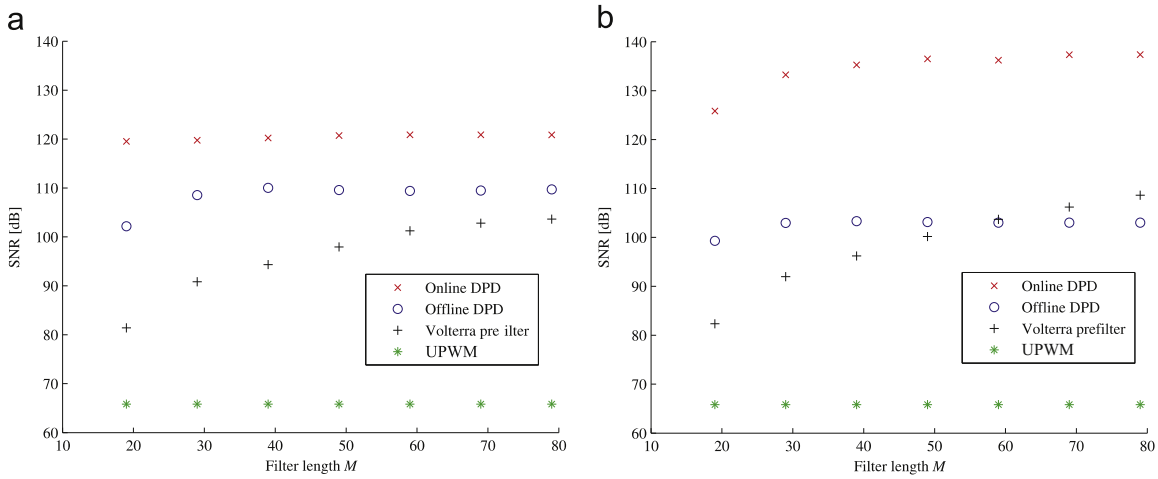

Fig. 6. SNR for Signal B (bandpass noise) as a function of the FIR filters length M. (a) $K=3$, (b) $K=5$.

is low-pass filtered (second order FIR with parameters: Fpass: 0.023, Fstop: 0.45, Astop: 1×10^{-7}) noise with Gaussian distribution Signal B is obtained by band-pass filtering (85-order FIR with parameters: Fstop1: 0.001, Fpass1: 0.041, Fpass2: 0.06, Fstop2: 0.1, Astop: 1×10^{-5}) noise with Gaussian distribution. Signal C is a segment of music which is also used for test purposes in [19]. Finally, Signal D is Signal C interpolated at twice the sampling frequency.

For the offline DPD, Signal A was used to train the DPD. The identified parameters were then used with the DPD excited by Signals B, C and D. The signal-to-noise ratio (SNR) was used as the quantitative index to compare the DPD with other alternatives. The SNR was computed as

$$\mathrm{SNR}[\mathrm{dB}] = 10\log\left(\frac{\sum_{n=1}^{N} x_n^2}{\sum_{n=1}^{N}(x_n-y_n)^2}\right), \tag{22}$$

where x_n and y_n are the input and output signals according to block diagrams of Figs. 3 and 4 and N is the length of the signals discarding any edge-effect due to the transient response of the filters. The results of the DPD are compared with standard digital PWM methods (UPWM) and with the Volterra prefilter (pth-order inverse) [19]. For each simulation the order of the Volterra prefilter is the same as the order of the DPD.

4.2. Simulations

Fig. 6(a) and (b) shows the SNR [dB] as a function of the length M of the predistorter filters for Signal B. The simulation was carried out for polynomials up to order $K=3$ and $K=5$. The online DPD has the best performance followed by the offline DPD which has a better performance than the pth-order inverse.

The simulations show that no noticeable improvement is obtained for the *offline* DPD when the order of the nonlinearity of the DPD is increased from $K=3$ to $K=5$. On the other hand, the *online* DPD increases its performance for $K=5$. At this point, the Volterra prefilter improves its performance for high values of M. For M greater than 60 and $K=5$ in Fig. 6(b), the Volterra prefilter slightly outperforms the offline DPD.

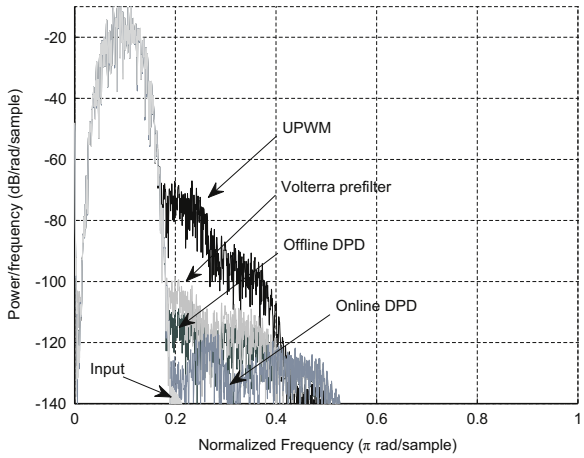

Fig. 7. Power spectral density of y_n using the different algorithms. For input Signal B, $M=39$ and $K=3$.

The power spectral density of the modulated signal used in Fig. 6(a) (input Signal B) for $M=39$ is shown in Fig. 7; the PSD of the outputs produced by each algorithm and the PSD of input signal are superimposed. Clearly, the standard UPWM method introduces a great amount of distortion, even for this low-frequency signal. For a signal with high frequency content the distortion will be even greater for the UPWM approach. The reduction of the spectral regrowth produced with each of the algorithms can be noticed in the figure.

Fig. 8(a) and (b) shows the SNR [dB] for $K=3$ as a function of the length M of the predistorter filters for the segments of music (Signals C and D). The SNR is higher for the interpolated Signal D, a result that is in agreement with the analysis of the modulation nonlinearity performed in Section 2.1 since, as depicted in Fig. 2 the contribution of the nonlinear terms of the PWM is decreased

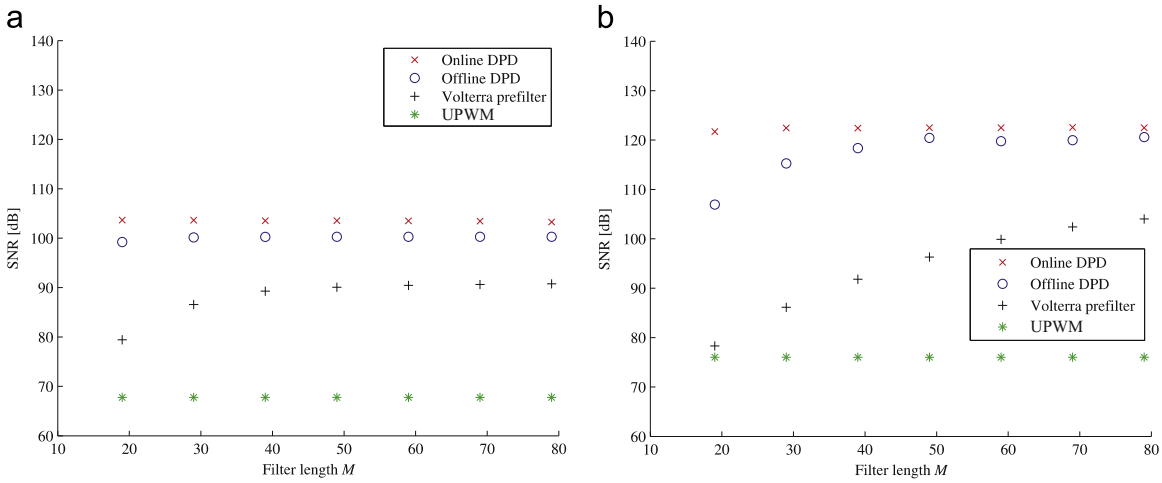

Fig. 8. SNR for Signals C and D (music signal without and with interpolation) as a function of the FIR filters length M with $K=3$. (a) $K=3$ and no interpolation (Signal C), (b) $K=3$ with interpolation (Signal D).

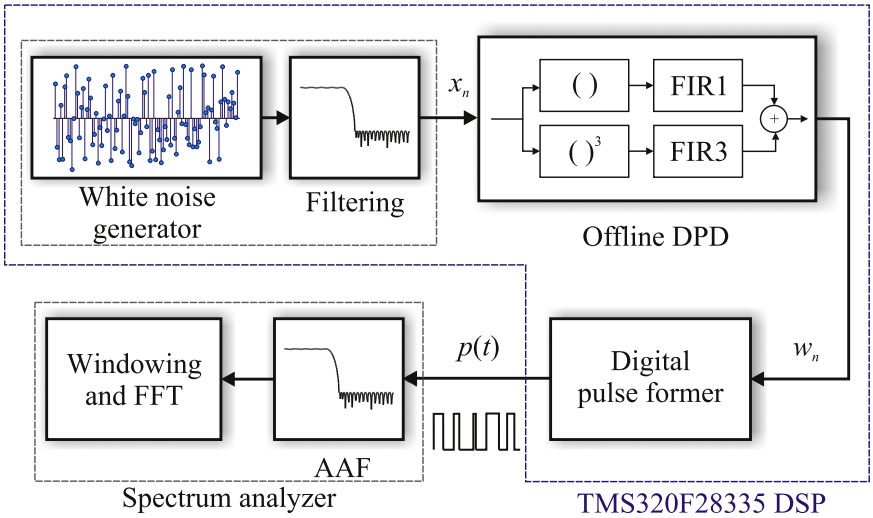

Fig. 9. Block diagram of the experimental setup.

when the signal bandwidth is reduced. The simulation with a musical signal shows the good performance of the DPD under a real-world signal which can actually drive a class-D switching amplifier in practice.

Simulations show that the DPD has a very good performance under typical finite bandwidth signals even when its parameters are computed offline. The improvement in the performance is even more remarkable under short filter length operation (low M), which is a desirable condition since the computational load and the delay introduced in the signal path are reduced. The DPD also shows a good performance using only the linear term and the cube factor $(x_n)^3$ achieving SNR values above 100 dB as shown in Fig. 7 for $M=39$. The offline DPD only requires to compute two length-39 FIR filters, one for x_n and other for $(x_n)^3$. This operating condition was verified experimentally.

4.3. Experimental measurements of the offline DPD

The proposed offline DPD with $M=39$ and $K=3$ was tested using a DSP development board (TMS320F28335). The offline DPD is computed in real-time at a frequency of 50 kHz which is equal to the carrier frequency (PWM frequency). The experimental setup is shown in the block diagram of Fig. 9. The input signal x_n is Signal B (bandpass noise), generated in real-time by filtering white Gaussian noise limiting its frequency content between 1.25 kHz and 8.75 kHz. Once x_n is processed with the offline DPD to obtain w_n, the actual PWM signal $p(t)$ is generated using an standard module of the DSP. The PWM signal was measured with a dynamic signal analyser (SR785) to obtain the frequency spectrum with and without the DPD. The results are shown in Fig. 10 in the frequency range 0 Hz to 25 kHz.

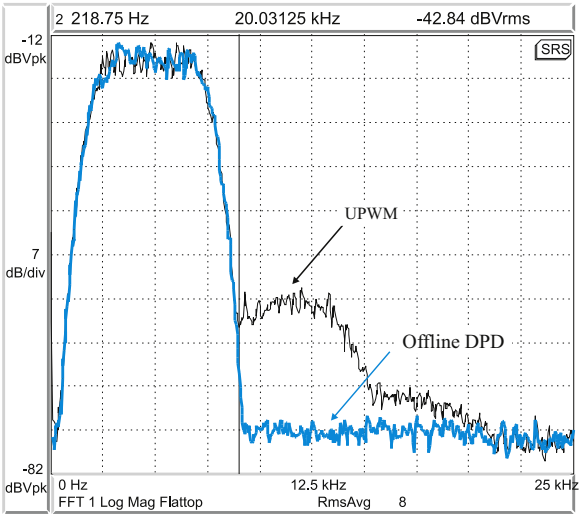

Fig. 10. Spectral measurement for input Signal B, $M=39$ and $K=3$.

The PWM without DPD (UPWM) shows distortion components between 8.75 kHz and 20 kHz (within the baseband) that are not present in the modulating signal, caused by the nonlinear behavior of UPWM. The DPD reduces these unwanted frequency components appearing between 8.75 kHz (3.5 divisions) and 20 kHz (8 divisions).

5. Conclusions

An adaptive predistorter (DPD) for digital PWM was developed. The proposed DPD reduces the distortion caused by the nonlinear behavior of the PWM modulation and is suitable for class-D switching amplifiers for digital modulating signals with finite bandwidth. The generalized Hammerstein model for PWM was reviewed and it was shown that for practical applications preserving up to the 7-th power of the model is enough to capture the nonlinear characteristic of the PWM in baseband. Simulations were performed for the DPD with online and offline parameter update. The results show that the SNR can be greatly increased as compared to standard PWM methods. The proposed DPD achieves satisfactory performance even when linear and cubic terms and short FIR filters are used. The behavior of the DPD with offline update was verified using a real-time implementation in a DSP showing the feasibility of the implementation for practical applications.

References

[1] A. Unsal, A.R. von Jouanne, V.L. Stonick, A DSP controlled resonant active filter for power conditioning in three-phase industrial power systems, Signal Process. 82 (11) (2002) 1743–1752.
[2] C. Pascual, Z. Song, P. Krein, D. Sarwate, P. Midya, W. Roeckner, High-fidelity PWM inverter for digital audio amplification: spectral analysis, real-time DSP implementation, and results, IEEE Trans. Power Electron. 18 (1) (2003) 473–485.
[3] K. Hausmair, S. Chi, P. Singerl, C. Vogel, Aliasing-free digital pulse-width modulation for burst-mode RF transmitters, IEEE Trans. Circuits Syst. I, Reg. Pap. 60 (2) (2013) 415–427.
[4] B.-H. Gwee, J.S. Chang, V. Adrian, A micropower low-distortion digital class-D amplifier based on an algorithmic pulsewidth modulator, IEEE Trans. Circuits Syst. I, Reg. Pap. 52 (10) (2005) 2007–2022.
[5] V. Adrian, J.S. Chang, B.-H. Gwee, A low-voltage micropower digital class-D amplifier modulator for hearing aids, IEEE Trans. Circuits Syst. I, Reg. Pap. 56 (2) (2009) 337–349.
[6] J. Noh, D. Lee, J.-G. Jo, C. Yoo, A class-D amplifier with pulse code modulated (PCM) digital input for digital hearing aid, IEEE J. Solid-State Circuits 48 (2) (2013) 465–472.
[7] H. Bresch, M. Streitenberger, W. Mathis, About the demodulation of PWM-signals with applications to audio amplifiers, in: IEEE International Symposium on Circuits and Systems, vol. 1, 1998, pp. 205–208.
[8] H. Black, Modulation Theory, Van Nostrand, New York, 1953.
[9] Z. Song, D.V. Sarwate, The frequency spectrum of pulse width modulated signals, Signal Process. 83 (10) (2003) 2227–2258.
[10] J. Goldberg, M. Sandler, Pseudo-natural pulse width modulation for high accuracy digital-to-analogue conversion, Electron. Lett. 27 (16) (1991) 1491–1492.
[11] J. Jin-Whi, M. Hawksford, An oversampled digital PWM linearization technique for digital to analog conversion, IEEE Trans. Circuits Syst. I, Reg. Pap. 51 (9) (2004) 1781–1789.
[12] G. Smecher, B. Champagne, Optimum crossing-point estimation of a sampled analog signal with a periodic carrier, Signal Process. 91 (8) (2011) 1951–1962.
[13] B. Logan Jr., Click modulation, AT&T Bell Lab. Tech. J. 63 (3) (1984) 401–423.
[14] L. Stefanazzi, A. Oliva, E. Paolini, Alias-free digital click modulator, IEEE Trans. Ind. Inform. 9 (2) (2013) 1074–1083.
[15] L. Ding, G. Zhou, D. Morgan, Z. Ma, J. Kenney, J. Kim, C. Giardina, A robust digital baseband predistorter constructed using memory polynomials, IEEE Trans. Commun. 52 (1) (2004) 159–165.
[16] D. Morgan, Z. Ma, J. Kim, M. Zierdt, J. Pastalan, A generalized memory polynomial model for digital predistortion of RF power amplifiers, IEEE Trans. Signal Process. 54 (10) (2006) 3852–3860.
[17] F.H. Gregorio, S. Werner, J. Cousseau, J. Figueroa, R. Wichman, Receiver-side nonlinearities mitigation using an extended iterative decision-based technique, Signal Process. 91 (8) (2011) 2042–2056.
[18] M.Y. Cheong, S. Werner, M. Bruno, J. Figueroa, J. Cousseau, R. Wichman, Adaptive piecewise linear predistorters for nonlinear power amplifiers with memory, IEEE Trans. Circuits Syst. I, Reg. Pap. 59 (7) (2012) 1519–1532.
[19] S.O. Aase, A prefilter equalizer for pulse width modulation, Signal Process. 92 (10) (2012) 2444–2453.
[20] C. Eun, E. Powers, A new Volterra predistorter based on the indirect learning architecture, IEEE Trans. Signal Process. 45 (1) (1997) 223–227.
[21] Y.H. Lim, Y.S. Cho, I.W. Cha, D.H. Youn, An adaptive nonlinear prefilter for compensation of distortion in nonlinear systems, IEEE Trans. Signal Process. 46 (6) (1998) 1726–1730.
[22] F. Kuech, W. Kellermann, Orthogonalized power filters for nonlinear acoustic echo cancellation, Signal Process.: Appl. Speech Audio Process. 86 (6) (2006) 1168–1181.
[23] J. Huang, K. Padmanabhan, O. Collins, The sampling theorem with constant amplitude variable width pulses, IEEE Trans. Circuits Syst. I, Reg. Pap. 58 (6) (2011) 1178–1190.
[24] F.J.I. Doyle, R.K. Pearson, B.A. Ogunnaike, Identification and Control Using Volterra Models, Springer Science & Business Media, London, England, 2002, http://dx.doi.org/10.1007/978-1-4471-0107-9.
[25] D. Psaltis, A. Sideris, A. Yamamura, A multilayered neural network controller, IEEE Control Syst. Mag. 8 (2) (1988) 17–21.
[26] M. Schetzen, The Volterra and Wiener Theories of Nonlinear Systems, John Wiley & Sons, New York, USA, 1980.
[27] R. Raich, G. Zhou, Orthogonal polynomials for complex gaussian processes, IEEE Trans. Signal Process. 52 (10) (2004) 2788–2797.
[28] S. Haykin, Adaptive Filter Theory, 4th ed., Prentice Hall, New Jersey, USA, 2002.
[29] Z. Cai, E.-W. Bai, Making parametric Hammerstein system identification a linear problem, Automatica 47 (8) (2011) 1806–1812.
[30] E.-W. Bai, K. Li, Convergence of the iterative algorithm for a general Hammerstein system identification, Automatica 46 (11) (2010) 1891–1896.
[31] R. Horn, C. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1990.
[32] R. Nowak, B. Van Veen, Invertibility of higher order moment matrices, IEEE Trans. Signal Process. 43 (3) (1995) 705–708.
[33] R. Horn, C. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1991.